Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 191 (2015) 2626 – 2630

Procedia
Social and Behavioral Sciences

WCES 2014

The use of Concordance for teaching Vocabulary: A data-driven learning approach

Enes Yılmaz* Adem Soruç

Fatih University, Faculty of Education, English Language Teaching Department,, Büyükçekmece 34500, İstanbul/TURKEY
Sakarya University, Faculty of Education, English Language Teaching Department, Hendek 54300, Sakarya/TURKEY

Abstract

Lexical knowledge and opportunities or tools for learning new vocabulary words have increasingly become important over the last four decades. But, using concordance as a new tool for this purpose has not gained much attention. The present research, therefore, aims to find out, if any, whether concordance as a reference tool has an impact on teaching or learning vocabulary in an EFL setting. As an experimental study, this research involved two groups of students conveniently. One was experimental group; the other control. While the former was exposed to electronic concordance program (COCA), the latter was controlled by using traditional vocabulary instruction. Pre/post-test results showed that the class receiving vocabulary through concordance performed much more than the control class which received traditional vocabulary instruction. And further analysis revealed that the performance was statistically significant. In addition, the interviews randomly held with the students also supported the quantitative results. Some implications are given and a variety of suggestions made for language teachers at the end.

Selection and peer-review under responsibility of the Organizing Committee of WCES 2014

Keywords: Corpora; concordance; data-driven learning; vocabulary teaching/learning

1. Introduction

After 1960's, when the structural linguistics was substituted by the generative linguistics, the first computer

* Corresponding author. Tel. +90-216 280 33 33
E-mail address: eyilmaz@fatih.edu.tr

Selection and peer-review under responsibility of the Organizing Committee of WCES 2014
doi:10.1016/j.sbspro.2015.04.400

corpus, also known as Brown Corpus, was created by Brown Corpus, W. Nelson Francis and Henry Kucera. But it did not galvanize linguists or teachers; in contrast, made them think the pioneering study as "a useless and foolhardy enterprise" (Francis, 1992, p. 28) because it was the years when native speaker intuitions were highlighted, from corpora which were believed to be insufficient to obtain.

However, after years when corpora were found a very useful resource, this time the use of corpora and concordance, one of the corpus tools, has mushroomed. In other words, corpora have "revolutionized" (O'Keeffe, McCarthy & Carter, 2007, p. 21) dictionary making processes in language teaching. This revolution has also galvanized English language teachers, thus leading them to investigate the importance or the role that corpora plays in 1) finding out word frequency counts, 2) keyword analysis, 3) cluster analysis, 4) lexico-grammatical profiles, 5) lexicography, 6) grammar, 7) stylistics, 8) translation and 9) concordance (ibid.: 2007). It is concordance and its influence on teaching vocabulary that was explored in the present study.

2. Corpus and Concordance

O'Keeffe et al. (2007) describes corpus as a tool in order to collect either written or spoken, or electronic, texts and to store them on a computer for the later qualitative and quantitative use. This yields lots of benefits for vocabulary studies or grammatical studies of specific linguistic constructions. For instance corpora provides 'a detailed study of a particular grammatical construction that yields linguistic information on the construction, such as the various forms it has, its overall frequency, the particular contexts in which it occurs … and its communicative potential.' (Meyer, 2004, pp. 11-12)

To date a lot of corpus-based research has been conducted on different grammatical structures and their various uses such as appositives in English (Meyer, 1992); clefts and pseudo-clefts (Collins, 1991); infinitival complement clauses (Mair, 1990); past and perfective verb forms in various periods of English (Elsness, 1997); the modals *can/may* and *shall/will* in early American English (Kytö, 1991); and negation (Tottie, 1991).

Specifically for instance Coates (1983) analyzed different uses of modal verbs in writing and speech by combining two corpora with 1,725,000 words in total. Further came after a decade from Mindt (1995), who received considerable benefits from a larger group of corpora with almost 80 million words to do the same analysis on modal verbs in writing and speech. The results of both studies were similar despite the different number of words in their corpora.

Central to the entire discipline of corpora is the concept of concordance, which is the most popular text analysis program. O'Keeffe et al. (2007) describes it as a way 'to find every occurrence of a particular word or phrase' (p. 8). In fact this is what text analysts painstakingly did for many years '...in the Christian Bible…finding and recording every example of certain words' (O'Keeffe et al. 2007, p.8). However, it was manually concordanced. The past thirty years have seen increasingly rapid advances in the field of technology and thus concordancing programs. Today, thanks to these developments we can easily reach a deep insight as to how collocations of the word 'bargain' work together (see figure 1) such as '*strike a bargain*' (O'Keeffe et al. 2007). Concordancing programs (e.g. KWIC-key word in context) further produce basic sequences of words through concordance lines and sort out files as in the figure 1 below, thereby helping the text analysts specify or search for particular tags (Meyer, 2004).

Figure 1. A sample of concordance lines for "bargain"

1 blic-sector unions have been allowed to **bargain** away jobs for pay. In a deal
2 over ... The chancellor also asks us to **bargain** away whatever obligations or int
3 :your loss is Southampton's gain. A **bargain** buy at pounds 1 million this sea
4 weapons; and that the Russians will not **bargain** for cuts in something that Labou
5 in his shirt front. Scurra has struck a **bargain**,' he called out as he bustled fu

(Cambridge International Corpus)

3. Data-driven learning

Recent developments in the field of Computer Assisted Language Learning (CALL) have led to a renewed interest in language teachers and corpora users, who can now employ CALL to give learners hands-on experience or real data through concordance lines. As a result of this, learners receiving instruction on the use of concordance may get help from corpora itself in order to learn English, which is called "data-driven learning" (Johns, 1994; Hadley, 1997). Johns (2002) further describes it as a methodology through which teachers can 'make the learner a linguistic

researcher' (p. 108) in vast amount of data or a "Sherlock Holmes" (p. 108) who can find real examples of the language usage.

Learners can make psycholinguistic guessing games thanks to concordance before and whilst learning vocabularies or various grammatical structures. Put it simply, they can construct some hypotheses for the word that they search for and then throughout the concordance lines are able to confirm or disconfirm their hypotheses. Kettemann (1995) had stated that a concordance program can also be used as "an extremely powerful hypothesis testing device." (p. 4)

Furthermore, language teachers or learners themselves can use concordance as a "language-learning activity" (Gavioli 1997, p. 84), or language learning tool with a vast amount of information from which learners can 'conduct inductive explorations of grammatical constructions' (Meyer, 2004, p. 27). In other words, they can gain an 'ability to see patterning in the target language and to form generalizations' (Johns, 1991, p. 2) in the real language usage.

The issue of whether corpora or concordance is effective at teaching vocabulary has been a controversial and much disputed subject in recent years. Thus, the extent to which concordance can provide a data-driven learning opportunity to learn vocabulary is the main impetus behind the present research. Put simply, the research questions were addressed as follows:

a) Is there a statistically significant difference between concordance (data-driven learning) group and control group which received traditional vocabulary instruction?
b) What do the participants in the concordance group think about the use of concordance in learning vocabulary?

4. Methodology

4.1. Research Design

A non-equivalent (NEC) experimental research design (Lynch, 1996) was used to conduct the study. It is not a true experimental study because the participants were not randomly selected for the study, although it involved a program (concordance) group and a control group. Data were collected thanks to quantitative and qualitative instruments. Put it another way, the participants took a pretest before the study, and a posttest after the implementation of the study (Research Question 1) and then five were randomly selected for the interview (Research Question 2).

4.2. Participants

Forty Turkish EFL students, in two classes at a private English language education center in Istanbul, Turkey, were conveniently recruited for the study in the fall semester of 2012-13 academic years. All of the participants were aged between 14 and 16. While just over half the sample (60%) was female; 40% male. The English proficiency level of the participants was pre-intermediate and it was homogeneous because they had to take a reliable proficiency exam before starting to get education at the center.

4.3. Data Collection Instruments and Procedure

Prior to the study, a word recognition test (pre-test) was given to all students, in which they indicated whether they knew the word by marking one of the columns as

a) *I know what this word means* and b) *I don't know what this word means*

If they marked the "a" option, this time in order to get more concrete outcomes, the participants also had to give the definition of the word either in his/her L1 (mother tongue) or in L2 (English). The aim of the pretest was twofold: it first aimed to eliminate those familiar words to the participants; second, to level out their vocabulary knowledge before the research. At the end left 23 target unfamiliar words to teach both in one concordance (program) group and one control group. Besides, the pre-test also included non-target words in order to disguise the target words.

Both groups took eight-class hour instruction from the first author. The program group used an electronic concordance program, known as *"Corpus of Contemporary American English (COCA),* to learn the target words inductively through concordance lines. They were given instructions on how to use COCA and search for the word within. The control group received definitions, synonyms and mother tongue equivalence of the words. Such vocabulary activities as "fill in the gaps, match the definition and cloze-tests" were performed.

After the instructions were completed, the participants this time received another posttest, which was exactly the same as the pretest.

To find out further, qualitative, support, five participants were randomly selected from the program group and interviewed. It was semi-structured interview with some questions asking about what they think or feel about concordance and through which vocabulary learning.

5. Results

The analysis was not conducted between pre and posttest scores because all the participants' knowledge was leveled out before the instructions started after the pretest analysis. In other words, they were already exposed to the words that were not familiar to them during the instructions. Therefore an independent sample *t* – test was conducted to compare only the posttest scores for the concordance and control group. The *t* – test analysis found a statistically significant difference in scores for the concordance group (M = 84. 50, SD = 10.659) and the control group, M = 70. 00, SD = 10; *t* (38) = 2.951, *p* = .016 (two-tailed). The magnitude of the differences in the means was very large (eta squared = 0.18).

The quantitative results have revealed that both instructional groups enhanced their performance on the word recognition test but it was greater for the concordance (experimental) group. The traditional vocabulary instruction group could not perform quite as good as the concordance group when recognizing the target words. Although those in the control did activities (e.g. fill in the blanks, synonym finding etc.), they were not able to recognize as many words as those in the concordance group, who searched for the word through concordance lines and learned them inductively. And the difference was statistically significant. Thus, the first research question was addressed.

To be able to find a qualitative support, interviews were held with randomly selected five students in the concordance group. Their responses were recorded and transcribed for the pattern coding. The analysis showed that concordance to search for the words and learn inductively was preferred by the participants. So the second research question was addressed. One stated

> I liked these lessons, because I personally used computer and it was an enjoyable way for learn vocabulary. The most enjoyable part was searching words in different colors in that search program (COCA).

Another stated

> My teachers want me to write same words many times with their Turkish meanings and I hate it. But, COCA is very good. I learn words myself. I see them in lines.

Another explained

> The words in lines are repeating which is good for me. I can notice the words before and after the lines.

6. Discussion and Conclusion

This experimental research study aimed to investigate the role of the concordance on teaching or learning vocabulary through an inductive approach. It was also controlled. The results showed that concordance had greater effect on teaching vocabulary than the traditional vocabulary instruction. It was effective, because the students did search for the target words and learned by themselves, which provided data-driven learning. As Johns (2002) stated, the students became a linguistic researcher investigating the real meaning of the target word in a real context. As they got used to concordance as a tool, it started to become a enjoyable language learning activity for the students (Gavioli, 1997).

The students were able to make inductive explorations (Meyer, 2004) through concordance and thus became

autonomous learners. More than that, they made psycholinguistic guessing games while searching the words through the lines. They used the program as a "hypothesis testing device." (Kettemann, 1995, p. 4)

The qualitative results also confirmed the quantitative ones. The students stated that they enjoyed using the program. They found the words in repetitive context through the lines, which provided them to learn vocabularies incidentally and inductively.

In conclusion, this study revealed the greater effect of concordance in teaching English vocabulary than the traditional way. When considering the results, the small number of the participants, selection of them, and the use of only word recognition test should be taken into account. And many more studies with a large number of participants should be conducted to validate the positive effects of concordance and further the results of this study.

References

Coates, J. (1983). *The semantics of the modal auxiliaries*. London: Croom Helm.

Collins, P. (1991). *Cleft and pseudo-cleft constructions in English*. Andover: Routledge.

Elsness, J. (1997). *The perfect and the preterite in contemporary and earlier English*. Berlin and New York: Mouton de Gruyter.

Francis, W. N. (1992). Language corpora B.C. In J. Svartvik (Ed.), *Directions in corpus linguistics: proceedings of Nobel symposium*, 82 (pp. 17–32). Stockholm.

Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 83–99). London: Longman.

Hadley, G. (1997). Sensing the winds of change: An introduction to data-driven learning. http://web.bham.ac.uk/johnstf/winds.htm.

Johns T. (1991). From print out to handout: Grammar and vocabulary teaching in the context of data-driven learning, *CALL Austria, 10,* pp. 14-34.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on Pedagogical Grammar* (pp. 293–313). New York: Cambridge University Press.

Johns, T. (2002). Data-driven learning: the perpetual challenge. In B. Kettemann, and G, Marko (Eds.), *Teaching and Learning by Doing Corpus Linguistics* (pp. 107-117). Amsterdam: Rodopi.

Kettemann, B. (1995). On the use of concordancing in ELT. *TELL & CALL, 4*, pp. 4–15.

Kytö, M. (1991). *Variation and diachrony, with early American English in focus. Studies on 'can'/'may' and 'shall'/'will'*. University of Bamberg Studies in English Linguistics 28. Frankfurt am Main: Peter Lang.

Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. New York: Cambridge University Press.

Meyer, C. F. (1992). *Apposition in contemporary English*. Cambridge University Press.

Meyer, C. F. (2004). *English corpus linguistics: An introduction*. Cambridge University Press.

Mair, C. (1990). *Infinitival complement clauses in English*. Cambridge University Press.

Mindt, D. (1995). *An empirical grammar of the English verb*. Berlin: Cornelsen.

O'Keeffe, A., McCarthy, M., and Carter, R. (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.

Tottie, G. (1991). *Negation in English speech and writing. A study in variation*. San Diego: Academic Press.